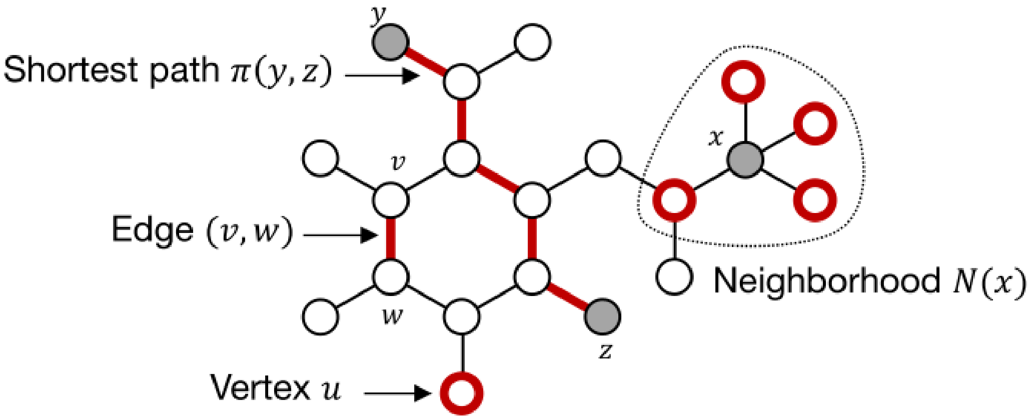
基本概念

图的基本概念



- **节点v的邻居数量：**$N(v)$
- **图上的有序游走序列w：**$w=(u,\ldots,v)$
- **图G与H同构：**$G\simeq H$

> 存在一个双射函数$\varphi:V(G)\to V(H)$，使得所有的节点对$(u,v)\in E(G)$有且仅有$(\varphi(u),\varphi(v))\in E(H)$

- **邻接矩阵A**
- **度矩阵D**
- **图拉普拉斯矩阵（Graph Laplacian）：**$L=D-A$
- **标准化图拉普拉斯矩阵（Normalized Graph Laplacian）：**$\hat{L}=D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$

> 定理：
>
> 1. 标准化图拉普拉斯矩阵是对称、半正定矩阵，特征值$\lambda_1,\lambda_2,\ldots,\lambda_n$满足$0\le\lambda_i\le 2$
> 2. 在度为d的规则图（即，图中每一个节点都有d个邻居）上，标准化图拉普拉斯矩阵、图拉普拉斯矩阵和邻接矩阵的特征向量一致
>
> 图拉普拉斯矩阵和标准化的图拉普拉斯矩阵可以作为线性算子作用于函数或者向量上，例如：
>
> $$\|f\|_L=<f,Lf>=\sum_{i,j}(f_i-f_j)^2 \tag{6}$$
>
> 以上起到了对图邻接节点的正则化作用（信号需相似）。

- **关联矩阵（Incidence matrix）M：**在$n\times n^2$的矩阵中，每个顶点-边对满足$m_{ue}=\{1\text{ iff }(u,v)\in E\}$

> $L=M\cdot M^T$，矩阵A,L,M包含相同的信息

核方法的基本概念

- **概念：**核方法指机器学习算法，其本质为学习衡量数据点对间相似度的评价方法—核（kernels）
- **核：**若函数$\kappa:\chi\times\chi\to\mathbb{R}$为核，则存在一个希尔伯特空间$\mathcal{H}_k$和特征映射$\phi:\chi\to\mathcal{H}_k$使得对所有的$x,y\in\chi$都有$\kappa(x,y)=<\phi(x),\phi(y)>$，其中$<\cdot,\cdot>$定义为空间$\mathcal{H}_k$上的内积。当且仅当$\kappa$是半正定函数，特征映射$\phi$存在

- **格拉姆矩阵（Gram matrix）$\mathbf{K}$**：核κ的格拉姆矩阵为对于矩阵中的每个元素K_{ij}都有$K_{ij}=\kappa(x_i,x_j)$

> 在核方法中，一般不需要对特征映射ϕ进行显式定义，通过对格拉姆矩阵的求解即可，例如：
>
> 高斯径向基函数（Gaussian radial basis function）：$\kappa_{\mathrm{RBF}}(x,y)=\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$，其中$\sigma$是带宽（bandwidth）参数

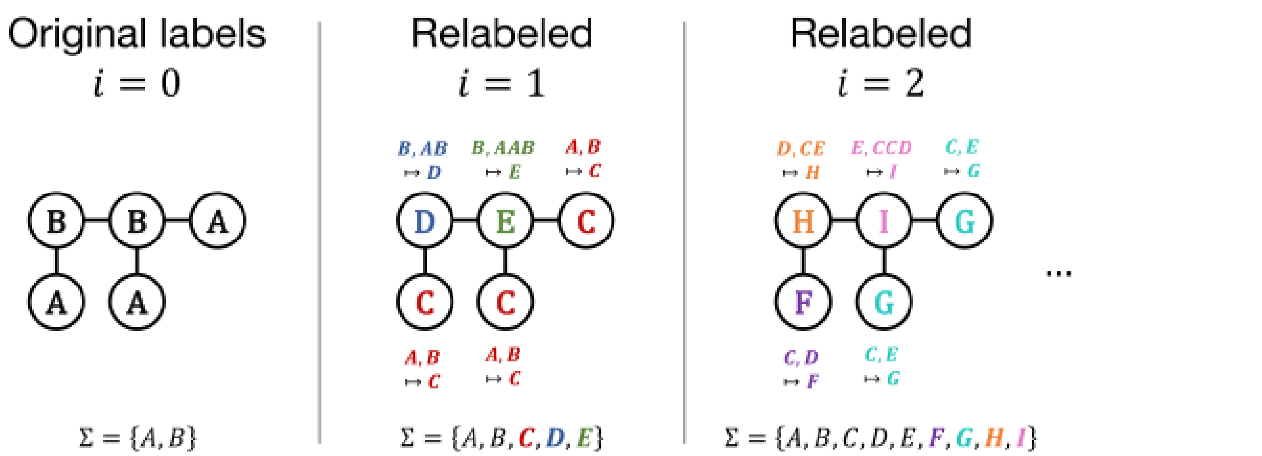
- **图核**：函数$\kappa:G\times G\rightarrow\mathbb{R}$
- **关系$\mathbf{R}$**：当且仅当$x_1,...,x_D$为x的组成部分时，$R(x_1,...,x_D,x)$在集合$X_1\times...\times X_D\times X$上为真，简洁表示为$R(\vec{x},x)$或$R(\vec{x})=x$，则$R^{-1}(x)=\{\vec{x}:R(\vec{x})=x\}$。假如$R^{-1}$是有限的，则称$R$对所有的$x\in X$是有限的
- **卷积核（Convolution kernels）**：适用于离散数据场景（字符串、树、图等），将数据分解为若干个组成部分（例如：图分解为节点或子图），然后结对判断组成部分的相似性，具体定义为：

> 令$\mathcal{R}=\mathcal{R}_1\times...\times\mathcal{R}_d$表示组件空间（space of components）可将目标$X\in\mathcal{X}$分解，令$R:\mathcal{R}\rightarrow\mathcal{X}$表示从组件到目标的映射，则当且仅当$x\in\mathcal{R}$组成目标$X\in\mathcal{X}$时，$R(x)=X$（这里的$x$等同于上面的$\vec{x}$）。则$R$-卷积核为：
>
> $$\kappa_{\mathrm{CV}}(X,Y)=\sum_{x\in R^{-1}(X)}\sum_{y\in R^{-1}(Y)}\underbrace{\prod_{i=1}^{d}\kappa_i(x_i,y_i)}_{\kappa(x,y)} \tag{7}$$
>
> 其中κ_i是定义在$\mathcal{R}_i$上的核，$\kappa(x,y)$可写为$\kappa(x,y)=\kappa_1\star...\star\kappa_d=\prod_{i=1}^{d}\kappa_i(x_i,y_i)$。
>
> 例，径向基核就是一种卷积核：
>
> $$\kappa(x,y)=\prod_{i=1}^{d}\exp\left(-\frac{(x_d-y_d)^2}{2\sigma^2}\right)=\exp\left(-\frac{\sum_{i=1}^{d}(x_d-y_d)^2}{2\sigma^2}\right)=\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \tag{8}$$

- **任意结构间的距离度量$d(x,y)$**：$d(x,y)=\sqrt{\kappa(x,x)-2\kappa(x,y)+\kappa(y,y)}$

威斯费勒-莱曼子树核（Weisfeiler-Lehman subtree kernel）

示例：

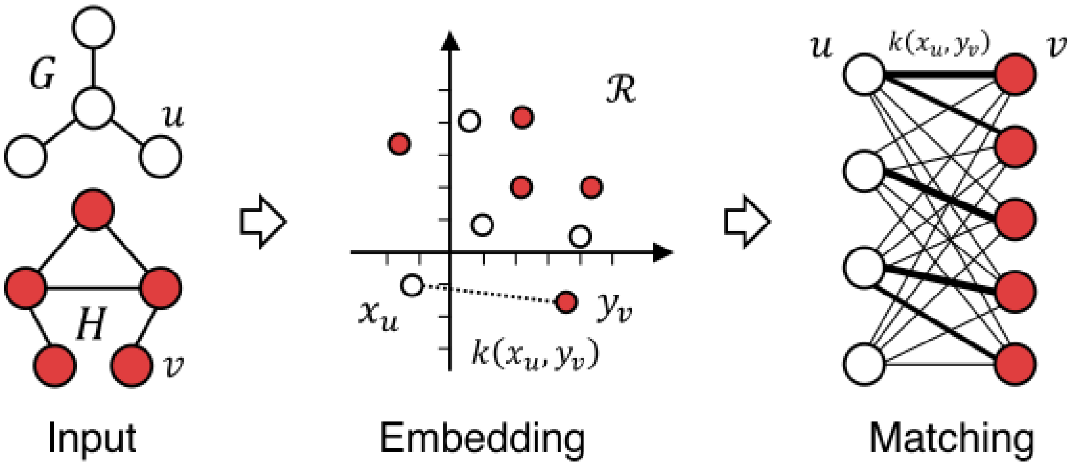


定义：令$l:V(G)\cup V(H)\rightarrow\Sigma$为观测节点的标签函数，通过$K$次循环迭代更新标签映射$l^i(v)=\mathrm{relabel}(l^{i-1}(v),\mathrm{sort}(\{l^{i-1}(u)|u\in N(v)\}))$。当$l^{i-1}$与$l^i$的像的基数一致时，迭代终止，即得到映射函数$\phi(G)$与$\phi(H)$。假如$G$和$H$有相同数量的同标签节点，则图同构。

> 1-WL核与标准GNN的思想一致，只是替换成了连续特征向量和网络层（迭代）

最佳指派（Optimal Assignment-OA）核

示例：



定义：令$X = \{x_1, \dots, x_n\}$和$Y = \{y_1, \dots, y_n\}$是源自组件空间$\mathcal{R}$的组件集合，且$\kappa : \mathcal{R} \times \mathcal{R} \to \mathbb{R}$是关于组件的基本核。则最优指派核为：

$$\kappa_A(X, Y) = \max_{\pi \in \Pi_n} \sum_{i=1}^{n} \kappa(x_i, y_{\pi(i)}) \tag{9}$$

其中Π_n为所有可能的$\{1, \dots, n\}$的置换。

> 相比与WL核，OA核仅选取最佳的匹配，在分类任务上的效果更好

子图结构相关核

顶点标签核（vertex label kernel）与边标签核（edge label kernel）

定义：$\kappa_{\mathrm{VL}} = \sum_{u \in V(G)} \sum_{v \in V(G)} \kappa(l(u), l(v))$（将顶点换成边即转换为边标签核）

图基元核（Graphlet kernel）

定义：令$\phi(G)_{\sigma_i}$为第i种同构类型σ_i的实例数量，则特征映射可以表示为：
$\phi_{\mathrm{GR}}(G) = \{\phi(G)_{\sigma_1}, \dots, \phi(G)_{\sigma_N}\}$。图基元核可定义为

$$\kappa_{\mathrm{GR}}(G) =< \phi_{\mathrm{GR}}(G), \phi_{\mathrm{GR}}(H) > \tag{10}$$